Learning Go/NoGo Terrain Classification

Abstract - **This paper presents a software system for image-based terrain classification that mimics a human supervisor's segmentation and classification of training images into "Go" and "NoGo" regions. The system identifies a set of image chips in the training images that span the range of terrain appearance. It then uses these exemplars to segment novel images and assign fuzzy Go/NoGo classification. System parameters adapt to new inputs, providing a mechanism for learning.**

Index Terms – terrain classification, computer vision, machine learning, exemplar memory

I. Introduction

Unstructured vision-based navigation continues to be an especially difficult problem for small robotic systems. If they are even equipped with a vision system, monocular and stereovision video remain the systems of choice for small inexpensive robots. In this paper, we present an approach to automated image segmentation and terrain classification using exemplars, or small image samples, to represent the variety of terrain appearance.

Exemplars are used as cluster seeds to segment the terrain. Local pieces of terrain are assigned to the exemplar to which they are most similar in appearance. The pieces of terrain then inherit the terrain class membership of the exemplar. Exemplar models assume that intact stimuli are stored in memory, and that classification or recognition is determined by the degree of similarity between a stimulus and the stored exemplars. Simple generalization effects explain correct classification of novel (previously unseen) instances of categories. Only the item information is used for classification decisions, and that categorization relies on the comparison of a new stimulus with known exemplars of the category.

Exemplar models are the most parsimonious models of categorization in terms of the underlying associative mechanism [1]. Exemplar based learning was originally proposed as a model of human learning in Ref. [2], and has since been shown to explain both human and animal visual classification performance significantly better than alternative hypotheses of feature-based and prototype-based processing [3,4].

Various researchers have begun to develop methods to forecast traversability using estimates of geometrical properties inferred from non-contract sensors. References [5] and [6] developed a fuzzy-rule-based system to mimic human "high/medium/low" trafficability assessment based on measures of roughness, slope and distance between obstacles computed from stereo imagery. The system was targeted for planetary rover environments. Reference [7] used a stereo color vision system together with a single axis LADAR to classify terrestrial terrain cover and detect obstacles. They noted that the color-based classification system could be made more robust by considering texture of regions and shape features of objects. Reference [8] defined a trafficability index equal to the weighted sum of the slope and roughness estimated from line-scanning laser rangefinder data. Reference [9] classified terrain as impassible (NoGo) if any of several properties were above a threshold: height variation, the surface normal orientation, and the presence of an elevation discontinuity (all estimated from LADAR imagery). Reference [10] developed a rule-based system for terrain classification from LADAR and color camera imagery.

Appearance based approaches do not attempt to directly estimate geometrical properties and then infer traversability. Instead, they associate the operator's assessment of trafficability directly from the terrain appearance. The operator's trafficability assessment is not restricted to geometrical properties, but can also reflect surface properties (e.g., friction, resistance, sinkage) and factors that do not affect traversability but which nonetheless exclude certain terrain (e.g., the risk of being run over by a car or the need to avoid detection by staying in shaded areas).

Various applications could benefit from automatic methods to segment and classify terrain from images, such as virtual reality simulated terrain, mobile robot navigation, combat engineering planning, and land cover analysis for ecological studies. These applications address different scales, terrain features and classes of interest. It is unlikely that any specific segmentation and classification criteria would be suitable for all of these applications. Nonetheless, the applications have important similarities. In all cases, we implicitly assume that local areas with similar appearance should be grouped together in any segmentation, and that they are likely to be representatives of the same terrain class. We also implicitly assume that we know in advance what terrain classes we are interested in and what they commonly look like. For the purposes of this research, we assume that the segmented terrain regions or regions of the same terrain class do not have any a priori constraints on their geometric shape or global organization. We also assume that there are no a priori constraints regarding which terrain classes can be adjacent to each other.

The approach is currently implemented as a software system designed to provide considerable flexibility in the choices of perspective transformation, resolution, scale,

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
15 APR 2006	**Journal Article**	**15-05-2006 to 15-05-2006**

4. TITLE AND SUBTITLE **Learning Go/NoGo Terrain Classification**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Robert Karlsen**	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army TARDEC ,6501 E.11 Mile Rd,Warren,MI,48397-5000**	8. PERFORMING ORGANIZATION REPORT NUMBER **#15264**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **U.S. Army TARDEC, 6501 E.11 Mile Rd, Warren, MI, 48397-5000**	10. SPONSOR/MONITOR'S ACRONYM(S) **TARDEC** 11. SPONSOR/MONITOR'S REPORT NUMBER(S) **#15264**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Submitted to ICRA Conference 2006.

14. ABSTRACT
This paper presents a software system for imagebased terrain classification that mimics a human supervisor?s segmentation and classification of training images into ?Go? and ?NoGo? regions. The system identifies a set of image chips in the training images that span the range of terrain appearance. It then uses these exemplars to segment novel images and assign fuzzy Go/NoGo classification. System parameters adapt to new inputs, providing a mechanism for learning.

15. SUBJECT TERMS
terrain classification, computer vision, machine learning, exemplar memory

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**6**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Fig. 1 Input training image and classification.

sampling and difference metric. In general, different choices will be appropriate for different applications. The software automatically builds a characteristic "basis set" of exemplars from training images. It provides an option for building a set of exemplars for each terrain class, with the union over the terrain classes being the basis set exemplars for an application. A second option is to build a set of terrain segmentation exemplars independent of the terrain classes, and then associate the exemplars with terrain classes. In its present form, the software does not attempt to resolve ambiguities when an area does not resemble any of the a priori terrain classes, or areas that have partial membership in two or more terrain classes. Instead, it produces a fuzzy classification, i.e., a segment of terrain can have partial membership in different terrain classes, and may be partially unclassified.

II. Technical Approach

The code is organized into two routines: one for training and one to apply segmentation and classification. At the end of training, the exemplar bank and associated data are stored in a file to be loaded before applying the segmentation and classification.

A. Training Images and Overlays

The user must provide a set of representative training images. Ideally, the training images would be drawn from the same distribution as the downstream application images. In practice, it may not be possible to ensure that the two image sets are drawn from the same distribution. The effect on segmentation and classification performance of different terrain, foliage, season, lighting, and weather between the training image set and test/application image set is a question for empirical investigation. In principle, the images can be multi-spectral with an arbitrary number of planes. The current code requires that the images be RGB or monochrome images stored in a standard image format.

For each training image, a corresponding terrain classification overlay is required. The overlay denotes which locations correspond to which terrain class. One approach is to use an N plane image, where N is the number of terrain classes and each plane is a binary image. An alternative approach is to use a single plane image, using integer values from 1 to N (for the N terrain classes), and zero for unclassified locations. This representation is more appropriate when there are a large number of terrain classes, or when the terrain classes constitute an ordered set, e.g., ordered by traverse ability cost or by speed-made-good. For purposes of demonstration, we use two terrain classes (e.g., "Go" and "NoGo" regions) and the overlays are stored as three-plane RGB images (the third plane is not used). The terrain classification is displayed as an RGB image in which one terrain class is coded red and the other is coded green, with blue used to code unclassified regions. An example of this is shown in Fig. 1, where the gravel driveway is designated as a "Go" region and everything else is designated a "NoGo" region.

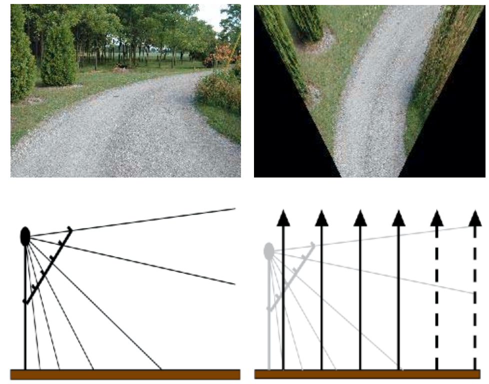


Fig 2 Camera image view and pseudo plan view.

B. Perspective Transformation, Resolution, Scale and Sampling

In some cases, a transformation from original camera perspective may be appropriate. In the camera image view, pixels represent the same angle (assuming lens distortion effects are minimal), but do not project onto equal areas of ground. Assuming the elevation of the camera is large relative to the variation in ground elevation in the scene, the pseudo plan view projection can be used to create a new image in which each pixel corresponds to the same ground area (see Fig. 2). The pseudo plan view projection is good for areas where the variation in elevation is small relative to the elevation of the camera, but produces distortion when this is not the case. An alternative projection is to restrict analysis to horizontal sub-bands within the image. The band view does not distort vertical objects, but retains the perspective distortion of the original camera image for flat earth regions.

Both the pseudo plan view and camera view options are supported in the current code. Both transformations require the size of the camera image, and the angle subtended by an individual pixel (we assume square pixels). The pseudo plan view projection requires three additional inputs: (1) the height of the camera above ground plane, (2) the distance on the ground from the spot below the camera to the ground projection of the bottom row of the image, and (3) the desired resolution of the projected image, i.e. the pixel width of the output projection in centimeters.

The camera band view also requires three additional inputs: (1) the image row number of the top row of the band, (2) the image row number of the bottom row of the band, and (3) the resolution for the band-view image (the angle of pixels in the band view image must be less than or equal to the pixel angle of the original camera image).

The user must also specify the analysis scale for terrain segmentation and classification. The segmentation and classification is based on exemplar image chips (square chips in the current code). The scale is the width of the exemplar chips. Membership in a terrain class is considered to be a bulk property of a local region, not a point-location property. The user must also specify the center-to-center spacing, or sampling distance, for the output segmentation and classification images.

C. *Image Space Transformation*

The purpose of the image space transformation is to amplify the importance of selected image properties. For example, the imagery can be transformed into a variety of color spaces. The importance of color could be strengthened or weakened by weighting different image planes. In addition to the RGB color coordinate system, we have experimented with the HSV (hue, saturation, value) system.

Constructing a multi-resolution pyramid representation and then applying weights to the image planes would allow the adjustment of high spatial frequency content relative to low spatial frequency content.

The space transformation could increase the dimensionality of the image space. Consider a monocular image input. The image could be processed through a bank of N spatial filters, such as edge and corner filters at different spatial scales and orientations. Each filter produces a single-plane output image.

D. *The Exemplar Basis Set*

The current code processes the training images one at a time. There is an option to find exemplars of each image independent of exemplars from other images, or to find only new exemplars sufficiently different from exemplars built from preceding images. The current image is chopped into chips at the specified scale and sampling distance. If the option was selected to process the image independently from previous images, all chips are nominated as potential exemplars. If the exemplar processing is in the context of previous exemplars, only chips whose minimum distance (in terms of the image metric) to existing exemplars is greater than the current clustering threshold are nominated as potential exemplars: chips that resemble current exemplars are not considered as possible new exemplars.

Each chip is compared to its neighbors within a specified radius to calculate the difference metric between it and each of its neighbors (the radius is a user input). The aggregate local difference between the chip and its neighbors is calculated as the weighted average of the mean and minimum differences (The weight is a user input. Weighting towards the minimum leads to a larger pool of exemplars, and weighting towards the mean leads to a smaller pool of exemplars). Chips similar to their neighbors are preferred over those that are different.

The code calculates a clustering threshold equal to the weighted sum of the minimum and maximum local differences over all chips (The weight is a user input. Weighting towards the minimum leads to a larger pool of exemplars and tighter clusters. Weighting towards the maximum leads to a smaller pool of exemplars and broader clusters). This threshold provides the system's adaptation ability. Training images with significant variability provide coarser segmentation over training images with lower variability, for the same size of exemplar bank.

Exemplars for the current image are selected iteratively. Initially, no chips are rejected. Of the non-rejected chips, the one with the minimum local difference is added to the bank of exemplars. All chips with difference less than the clustering threshold from the exemplar are rejected. This process is iterated until all chips have either been added to the exemplar bank or rejected. The exemplars for the current image are then merged with the bank of exemplars from the previous images.

E. *Image Chip Difference Metric*

Image difference metrics remain an open issue in the evaluation of image compression schemes. While it is easy to measure the amount of compression, and the encoding/decoding time, it is not clear how to measure the quality of the reconstructed image, i.e., its difference in appearance from the original. Different image characteristics are important depending on the image content, the questions at hand, and who is looking at the image.

Similarly, there is no obviously correct metric for measuring the difference between two images. Before the images are chopped into chips, they can be processed to balance the relevant image characteristics (see II.C Image Space Transformation). In principle, therefore, simple measures of the aggregate difference are all that are needed. Even so, there are many different ways to calculate the difference between two image chips, e.g.,

(1) the sum over all pixel locations and all image planes of the absolute value of the difference between the two images;

(2) the root sum square over all pixel locations and all image planes of the difference between the two images;

(3) the maximum over all image planes of the sum over all pixel locations of the absolute value of the difference between the two images;

(4) the sum over all pixel locations of the maximum over all image planes of the absolute value of the difference between the two images;

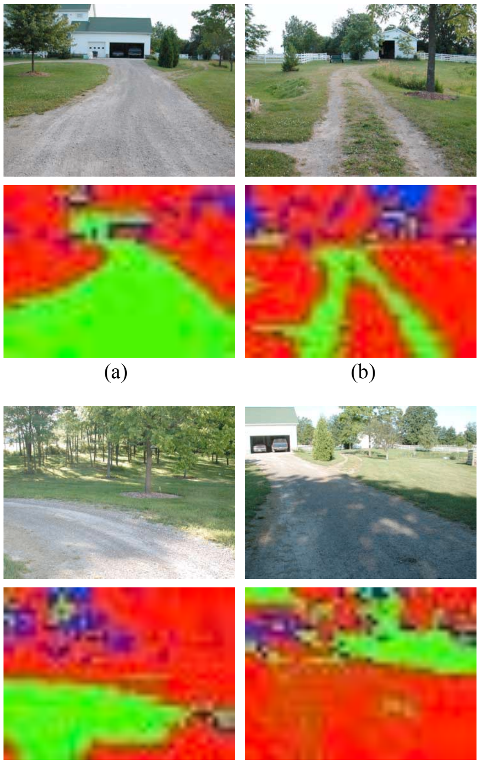
(5) the root sum square over all image planes of the difference in the mean values (over pixel locations) of the two images; and

(6) the root sum square over all image planes of the difference in the mean values and difference in standard deviations (over pixel locations) of the two images.

Two important classes of metrics are those computed from the difference between the images (metrics 1 through 4), and those computed from the difference in statistics computed from the individual images (metrics 5 and 6). While the code is set up to incorporate different metrics, all of the results in this paper used metric (1).

F. *Exemplar Membership in Terrain Classes*

Each image chip maps to a region in the terrain classification overlay. The terrain classification of the image chip is simply the expected membership in each of the terrain classes. It is possible that a chip could straddle more than one terrain class, or could straddle an unclassified portion of the



(a) (b)

(c) (d)

Fig. 3 Test images and resulting classification maps.

overlay. After the new exemplars are added to the exemplar bank, the current image is segmented using all of the exemplars in the bank. Each chip location in the image is assigned to the exemplar to which it is closest, provided the distance is less than the current clustering threshold. In some cases, some image chips may not be associated with any exemplar. For each exemplar in the bank, we accumulate the number of times the exemplar is "hit" by an image. The terrain class membership of the exemplar is the mean over all chips associated with the exemplar, of terrain class memberships of the chips. The terrain segmentation is converted to terrain classification by assigning each location the terrain class membership values of the exemplar associated with that image location.

G. *Output Illustration Controls*

The code contains options to output different images to illustrate and provide insight into the processing:

- the pseudo plan view or camera band view perspective transformation of the image;
- the pseudo plan view or camera band view perspective transformation of the terrain class overlay;
- the exemplar chips (at their location in the image) selected from the current image;
- the segmentation of the current image based on the current bank of exemplars; and
- the classification of the image based on the current bank of exemplars.

Fig. 4 Reconstruction of Figs. 3(a) and (d) using exemplars.

There is no obvious and correct way to represent the different segments for purposes of visualization. Color-coding shows the different segments, but does not give much insight into the basis for the segmentation. The code illustrates the segmentation in a way that provides direct visual insight into the basis for the segmentation. To visualize the segmentation, the code replaces each image chip with the exemplar chip that it is associated with (image chips not associated with any exemplar appear black) (See Fig. 4). When the sampling distance is less than the exemplar scale, the exemplars are blended in the reconstruction. The visualization image is the same size as the pseudo plan view or camera band view perspective image, so it is easy to directly compare the two. By using the exemplar chips themselves, the visualization image shows what the exemplars look like, and which image chips they are associated with. Finally, comparing the visualization to the perspective image gives prima fascia evidence of the credibility of the segmentation.

H. *Application for Segmentation and Classification*

The application routine reads in the filter bank and associated data produced by the training routine. It segments and classifies the test images one at a time. No changes are made to the exemplar bank or associated data. After pseudo plan view or camera band view perspective processing, the test image is chopped into chips at the specified scale and sampling distance. Each image chip is assigned to the closest matching exemplar, providing the match is within the current clustering threshold, otherwise the chip is unassigned. This produces the segmentation by exemplars. After the segmentation, each location is assigned the terrain class fuzzy membership of the segmenting exemplar. The classification image is at the resolution of the center-to-center sampling distance.

III. DEMONSTRATION RESULTS

This section illustrates the segmentation and classification system. The demonstration uses color-coding to show the terrain classification into Go (green), NoGo (red), and Unclassified (blue) regions. Fig. 3 shows classification results derived from the single training image in Fig. 1, where gravel is designated "Go" and everything else is "NoGo." Note the errors in (a) due to the building, in (c) due to the bright gravel patch, and in (d) due to the shadowed gravel. Fig. 4 shows an example of the reconstruction of images using the exemplar patches, as described in Sect. II.G. Adding a second training image (Fig. 5) to compensate for the misclassifications in Fig. 3 due to the shadowed gravel, results in the classification results of Fig. 6. Note the improvement to Fig. 6(d) compared

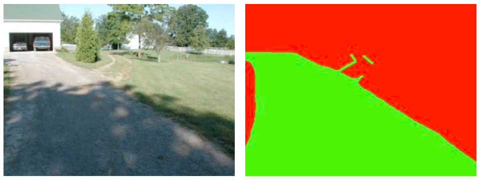


Fig. 5 Second training image and classification.

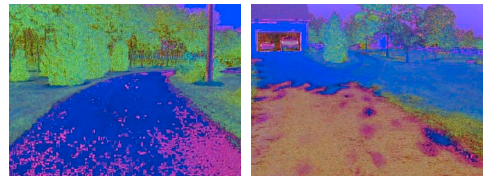


Fig. 7 HSV training images.

to Fig. 3(d). However, the overall classification map has become noisier.

To compensate for different lighting conditions, we turned to the HSV (hue, saturation, value) color coordinate system. Fig. 7 shows an RGB rendition of the two training images (Figs. 1 and 5) in the HSV color space. Fig. 8 shows classification results when using only the first training image. Note that the errors in Fig. 3, due to the house and the darkened gravel, are replaced by Unclassified regions. The addition of the second HSV training image results in much improved classification in Fig. 9. Fig. 10 shows that the classification results are only slightly degraded when just hue is used for classification.

However, hue alone is not sufficient in more complex scenarios. Fig. 11 shows the results of using just hue when the "Go" region includes both the gravel and grass regions in the training images. Fig. 12 demonstrates the improvement that is obtained when the other two dimensions (saturation and value) are also included.

IV. Findings and Observations

This paper has demonstrated an approach to image-based terrain segmentation and classification using exemplars. Exemplars provide a simple way to represent the characteristic color/luminance and spatial patterns of terrain. Since the exemplars are drawn from training images in such a way as to span the appearance of the training images, they are well suited to represent the variations of appearance without an a priori model of terrain appearance. The software system, as presented, allows for considerable flexibility to specify the perspective transformation, image space transformation, scale, resolution, sampling density, and image difference metric. Empirical research is needed to tune these options for specific applications. Preliminary results indicate the approach has potential to segment terrain in a manner that is consistent with subjective perception. The segmentation appears to be robust over changes in lighting, specific terrain, and automatic camera gain and contrast adjustments. Our preliminary results indicated that analysis in the camera band view was more useful for segmenting and classifying positive obstacles than the pseudo plan view. When presented with novel images, the camera band view was more likely to produce mixed Go/NoGo terrain classification, whereas the pseudo plan view was more likely to produce unclassified terrain segments. This may be due to the fact that the camera band view mixes different scales, whereas the pseudo plan view maintains more consistent scale.

The code performs quite well on the simplistic segmentation of gravel from other terrain. When presented with a combination of both grass and gravel, the system still performed reasonably well. Nonetheless, the preliminary analysis is not adequate to assess the value of this method of terrain classification for any specific application, e.g., robot navigation. More extensive testing, with a structured experimental objectives and design are needed to evaluate the applicability of this method of terrain classification for any specific application. The current code is reasonably fast, with the largest time consumption actually being the reconstruction of the segmentation images by inserting exemplars. But this step is for visualization purposes only. The method presented here does not address de-fuzzification, i.e., how to make discrete decisions based on the fuzzy membership, and does not address how to make discrete decisions when terrain class has partial membership in the "unclassified" set. The research

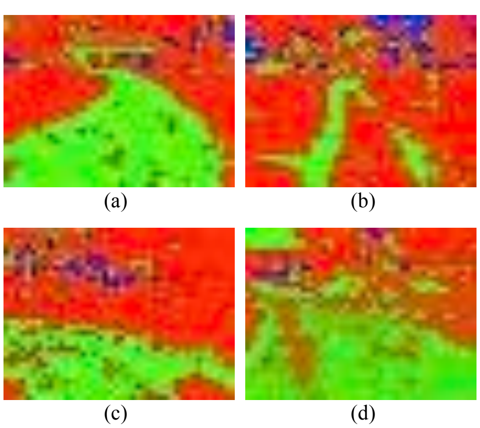


Fig. 6 Classification results with two training images.

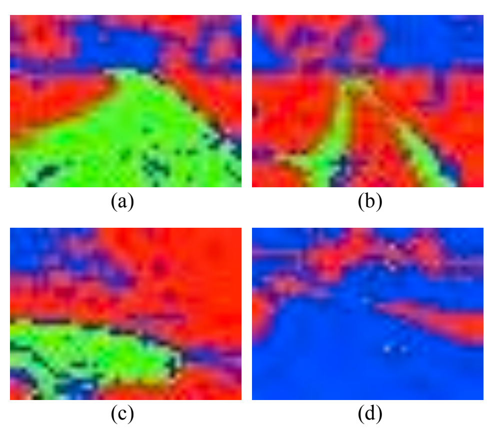


Fig. 8 Classification results with one HSV training image.

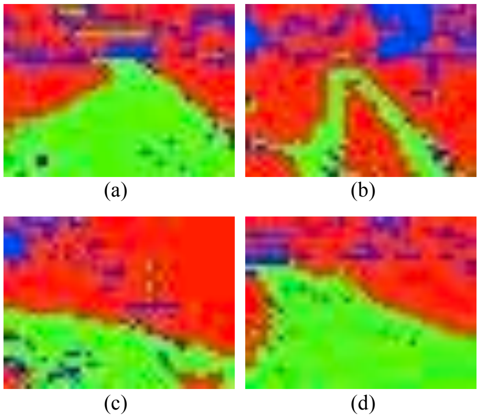


Fig. 9 Classification results with two HSV training images.

presented here does not address how to combine results obtained by analysis at different levels of resolution and/or scale. Further research in these topics is needed, in the context of specific applications.

Future work includes methods for pruning the exemplar bank, since the speed of the code is greatly influenced by the number of exemplars. The current code already prunes those exemplars that have not been used recently. But a more direct pruning method is also needed. We will explore a second training iteration that measures exemplar proximity and also an iteration that assesses the performance of each exemplar, keeping those that perform best in terms of classification. We also intend to investigate different color spaces, especially those that provide more uniform perceptual differences. Texture is known to be important and therefore, in future versions, we will add auxiliary image planes that explicitly include computed texture information. Since terrain appearance varies as a function of distance, we also anticipate fusing range data from a stereo camera system with the color and implicit texture information currently being used.

REFERENCES

[1] S. Chase and E.G. Heinemann, "Exemplar memory and discrimination," in *Avian Visual Cognition*, R.G. Cook, Ed., 2001.

[2] D. Medin and M. Schaffer, "Context theory of classification learning", Psychological Review, 85:3, 207-238, 1978.

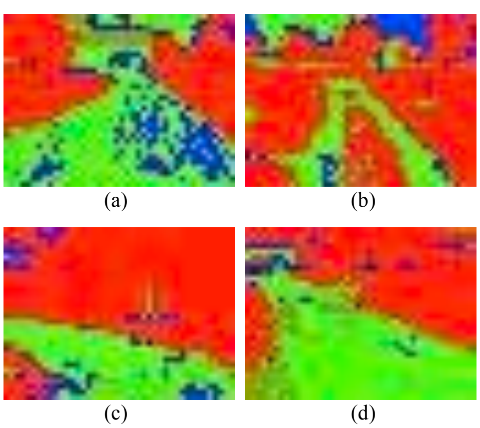


Fig. 10 Classification results using only hue.

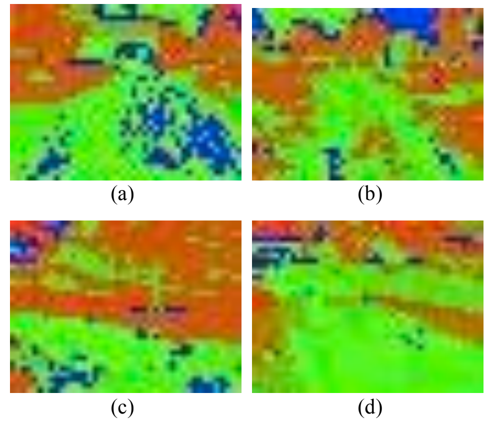


Fig. 11 Grass/gravel classification results using only hue.

[3] R.M. Nosofsky, "Tests of an exemplar model for relating perceptual classification and recognition memory," Journal of Experimental Psychology: Human Perceptual Performance, 17(1):3-27, 1991.

[4] C.W. Werner and G. Rehkamper, "Categorization of multidimensional geometrical figures by chickens (Gallus gaullus f. domestica): fit of basic assumptions from exemplar, feature and prototype theory," Animal Cognition, 4, 37-48, 2001.

[5] A. Howard, E. Tunstel, D. Edwards, and A. Carlson, "Enhancing fuzzy robot navigation systems by mimicking human visual perception of natural terrain traversability," Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf., Vancouver, B.C., Canada, pp. 7-12, July 2001.

[6] A. Howard, and H. Seraji, "Vision-based terrain characterization and traversability assessment," *J. Robotic Systems* 18(10), pp. 577-587, 2001.

[7] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Obstacle detection and terrain classification for autonomous off-road navigation," *Autonomous Robots* 18, pp. 81-102. 2005.

[8] C. Ye and J. Borenstein, "A method for mobile robot navigation on rough terrain," Proceedings of the IEEE International Conference on Robotics and Automation, pp. 3863-3869, 2004.

[9] D. Langer, J.K. Rosenblatt, and M. Hebert, "A behavior-based system for off-road navigation," IEEE Trans. On Robotics and Automation, Vol 10. No. 6, pp. 776-783, 1994.

[10] A. Sarwal, D. Simon, and V. Rajagopalan, "Terrain classification," Proc. SPIE Vol. 5083, Unmanned Ground Vehicle Technology V, 156-163, 2003.

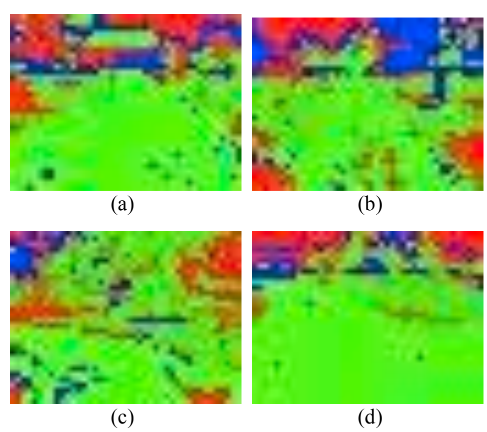


Fig. 12 Grass/gravel classification results with two HSV training images.